DOCUMENT RESUME

ED 068 333 SE 014 871

AUTHOR Funk, H. James
TITLE Test Construction Behavior of Pre-Service Secondary Science Teachers.
PUB DATE Apr 72
NOTE 19p.; Paper presented at the National Association for Research in Science Teaching Annual Meeting (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Educational Research; *Evaluation; *Preservice Education; Science Teachers; *Science Tests; *Secondary School Science; Teacher Attitudes; Test Construction; Tests
IDENTIFIERS Research Reports

ABSTRACT

Reported is a study of various aspects of test construction by science teachers. An attempt was made to examine: (1) variables affecting test-construction practice of pre-service science teachers; (2) attitudes of teachers as a result of instruction; and (3) relationship between test-construction behavior and attitudes. In addition, an attempt was made to compare testing practices of pre-service science teachers with another population of science teachers. The sample was drawn from secondary science majors enrolled in the science methods course and in student teaching. Two sample tests, obtained from pre-service science teachers, were analyzed and categorized according to the six levels of "Bloom's Taxonomy." The pre-, Post-, and delayed post-instructional data for determining subjects' attitudes toward the "taxonomy" were obtained by means of Word Association Scale (included in this report). The data were analyzed by factorial analysis methods. The results showed that tests prepared during student teaching contained higher proportions of knowledge level items than the tests prepared during the methods course. (Author/PS)

ED 068333

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Test Construction Behavior of Pre-service Secondary Science Teachers

H. James Funk
Indiana University at South Bend

Introduction

During the past two decades the emphasis on teaching science as an immutable body of knowledge has been supplanted by the notion that it ought to be taught as a process of acquiring knowledge. The recent curriculum developments that emphasize teaching science as inquiry reflect these changes in philosophy. Concommitant with these developments has been a resurgence of interest in teacher behavior and the effect this behavior has on the learner. Testing practices are an example of teaching behaviors that warrant closer examination. Research has provided evidence demonstrating the effect of questions on the learner. Increased learner achievement is one benefit which has been attributed to effective questioning. Numerous educators have stated that the levels of thought developed and exhibited by the learner are a function of the kinds of questions he encountered during his education. Questions have also been recognized as powerful motivational devices, influencing how the learner studies and the values he places on intellectual activities.

In spite of the curricular changes and the continued criticism of educators, questioning practices have not changed from those of 60 years ago. Teachers persist in

*Paper presented at the National Association for Research in Science Teaching (NARST) Annual Meeting, Chicago, Illinois, April, 1972.

FILMED FROM BEST AVAILABLE COPY

SE 014 871

asking questions requiring primarily memory skills. This use of predominantly low level questions has a limiting effect on the success of the new science curricula because it obscures the nature of science and leads the learner toward the mistaken impression that science is nothing more than a conglomeration of facts.

A plausible explanation for the failure of teachers to ask higher level questions is that they simply lack the necessary knowledge and skills. It is evident that developing the teachers' ability to use higher level questions is a necessary step toward achieving the broader goals of science education. Science methods courses should assume the task of preparing science teachers to meet the demands placed on them by the new curricula and the nature of inquiry in science. A means of accomplishing this end is to prepare teachers to use Bloom's Taxonomy as a model for questioning.

This study was designed to investigate the effectiveness of a method intended to develop improved test construction. The purpose of the study was to:

1. examine the variables affecting test-construction practices of pre-service science teachers
2. examine the effect of instruction on the attitudes of pre-service science teachers
3. examine the relationship between test-construction behavior and attitudes
4. determine whether the testing practices of pre-service science teachers were different from a population of science teachers.

The judgment of the desired improvement in testing practices was based on the premise that: (1) questions effect the attitudes of the learner; (2) questions to a large extent determine the cognitive skills the learner will develop; and (3) the possession of the intellectual skills is central to the understanding of science. Therefore, instruction was sought to promote a decrease in the use of questions requiring only memory skills.

The scope of the study was limited to secondary science majors attending Indiana University, Bloomington, Indiana. The study population consisted of secondary science majors who enrolled in the methods course during the period of the study. A subset of this population, those subsequently enrolling in student teaching, was selected for further study.

PROCEDURES

Test-Construction Data

Sample tests representing test-construction behavior were obtained during the methods course and during student teaching.

Sample Test 1 was prepared by all subjects as a required assignment for the methods course. The assignment, planning student evaluation, stated that the students were to prepare an objective-type test that:

1. could be used with a specific population of students that was to be described in detail,
2. contained a minimum of fifteen multiple-choice items and was at least twenty items in length, and
3. covered what, in their judgment, would be an appropriate unit of instruction.

The students were allowed to choose the content and resource materials they wished to use to complete the assignment. The students were not aware that the assignment would be used as data in a study.

One hundred and forty-nine tests were obtained during the methods course.

Sample Test 2 was prepared by the subjects while they were student teaching. Since the dispersal of student teachers throughout the state made individual contact impractical, letters of request were dispensed in pre-stamped, return-addressed envelopes containing the delayed post-instructional attitude measure form. In the request, the subjects were asked to prepare a test of a multiple-choice type for use in one of the classes he was teaching, and to return the test and completed form via mail. The area student teacher coordinators and supervising teachers were informed of the request and asked to encourage the student to complete the effort while he was student teaching.

Eighty-two of the methods students enrolled in student teaching during the period of the study. Attrition caused by a variety of factors reduced the return to 67 responses. Eighty-two percent return was judged to be an adequate sample.

Sample tests 1 and 2 were analyzed and items categorized according to the six levels of the Taxonomy.

Attitude Data

Observations on the subjects' attitudes toward the Taxonomy were obtained by means of a Word Association Scale. Pre, post, and delayed post-instructional observations were obtained.

Three sets of attitude data were collected. Delayed post-instructional observations only were collected from one group of subjects. Post and delayed post-instructional observations were collected from a second group. Pre and post-instructional data were collected from a third group.

Instrument Construction and Reliability

Attitudes. Data representing the subjects' attitudes toward the Taxonomy were obtained by their responses on the Word Association Scale. The scale initially consisted of 67 pairs of bipolar adjectives taken from Measurement of Meaning (Osgood, Suci, and Tannenbaum, 1957) and the stimulus concept "Bloom's Taxonomy." The scale was constructed in the following manner:

Concept: Bloom's Taxonomy

Polar term X 1: 2: 3: 4: 5: 6: 7 Polar term Y

1. Strongly associated with X
2. Moderately associated with X
3. Slightly associated with X
4. Associated with neither X nor Y
   or
   equally associated with X or Y
5. Slightly associated with Y
6. Moderately associated with Y
7. Strongly associated with Y

The subjects were instructed to identify the association existing between Bloom's Taxonomy and each response pair by marking the scale between each pair.

The scale was first administered to a group of methods students. The word pairs having no associative meaning (invoking primarily neutral responses) were deleted. The remaining pairs were submitted to a panel of judges familiar with the Taxonomy.

The judges were instructed to indicate the pairs having the closest relationship to the Taxonomy. From the pairs judged to have the closest association, fifteen pairs were selected for the final form. The form was comprised of six evaluative pairs, three potency pairs, and six oriented-activity pairs. The selected pairs were randomized with respect to polarity and dimension to construct the final form. The positive end of the scale was assigned a value of seven and the negative end of the scale was assigned a value of one with five equal intervals between the one and seven. The subjects attitude subscores were obtained by summing their responses in each of three dimensions.

Tests. Two hundred and sixteen tests were collected for the study. The items on each were examined and classified according to the six categories in The Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain (Bloom ed., 1956) as knowledge, comprehension, application, analysis, synthesis, and evaluation.

A random sample of ten tests was selected to check the investigator's reliability in using the Taxonomy. The intra-scorer reliability was checked using Scott's Formula for Scoring Reliability. Interscorer reliability was checked by establishing a panel of three judges, the investigator and two other judges familiar with the Taxonomy, and comparing their independent ratings of the items on the ten tests using an analysis of variance estimate of judge's ratings.

Analysis of Data and Results

The data were analyzed to answer the following research questions:

1. Do the students' grade point averages or the students' majors affect their test-construction behavior during the methods course and during student teaching?

2. Does cognitive instruction in the Taxonomy cause a change in attitude that persists during the student teaching assignment?

3. Do the subjects' attitudes toward the Taxonomy affect their test-construction behavior?

4. Do the subjects' test-construction behavior differ from the test-construction behavior of science teachers cited in an earlier study?

Question 1: To answer this question, the subjects were ranked on the basis of GPA and classified as high or low using a median split, categorized according to major (biology, general science, or physical science), and test-construction data were collected. The proportions of items occurring at the six levels of the Taxonomy were calculated for both sample tests. One criterion variable, proportion of knowledge level items, was examined.

A 2 x 2 x 3 factorial analysis of variance model with two levels of repeated measures on the criterion variable, two levels of GPA, and three levels of major was generated.

TABLE 1. ANALYSIS OF VARIANCE FOR MAIN AND INTERACTION EFFECTS ON SUBJECTS' USE OF KNOWLEDGE LEVEL ITEMS ATTRIBUTABLE TO TESTS, LEVEL OF GPA, AND MAJOR AREA OF PREPARATION

Sources of Variation	d.f.	s.s.	m.s.	F
Between Subjects	66	56,614.28		
1 Major (A)	2	8,517.10	4,258.55	5.68**
2 GPA Level (B)	1	2,237.91	2,237.91	2.99
12	2	126.98	63.49	.08
Subjects in 12	61	45,732.29	749.71	
Within Subjects	67	29,561.00		
3 Tests (C)	1	10,926.11	10,926.11	43.30***
13	2	1,294.68	647.34	2.57
23	1	176.88	176.88	.70
123	2	1,771.59	885.79	3.51*
Subjects x 3 in 12	61	15,391.74	252.32	

*$p < 0.05$

**$p < 0.01$

***$p < 0.001$

Table 1 shows:

1. a statistically significant difference among levels of major ($p < 0.01$). Further analysis by use of t-tests revealed no statistically significant differences between biology and general science means, however, statistically significant differences in mean proportions of knowledge level items were observed between

physical science and biology and physical science and general science (Table 2).

TABLE 2. t-TESTS ON MEAN PROPORTION KNOWLEDGE LEVEL ITEMS BETWEEN GENERAL SCIENCE AND BIOLOGY, BIOLOGY AND PHYSICAL SCIENCE, AND GENERAL SCIENCE AND PHYSICAL SCIENCE

N	General science mean	Biology mean	Physical Science mean	Mean gain	d.f.	t-value
122	60.65	63.06		-2.41	120	.510
92		63.06	34.58	-28.48	90	3.781***
54	60.65		34.58	-26.13	52	3.806***

***$p < 0.001$

2. no significant differences between levels of GPA.

3. a highly significant difference between levels of repeated measures on the criterion variable. The source of the difference was an increase of the mean proportion of knowledge level items from the methods course to student teaching (50.8 to 68.9).

4. a significant 123 interaction.

Question 2. Does instruction in the use of the Taxonomy cause a change in attitude that persists through the student teaching assignment?

Attitude data were collected from three groups of subjects. Group 1 data consisted of delayed post-instructional observations. Group 2 data consisted of post and delayed post-instructional observations. Group 3 data consisted of pre and post-instructional observations.

TABLE 3. t-TEST COMPARISON OF MEAN DELAYED POST-INSTRUCTIONAL ATTITUDE SCORES FOR GROUP 1 AND GROUP 2

	Subscore Variables			
	Evaluative	Potency	Oriented Activity	Total
Group 1 N=22	35.455	16.318	35.045	88.818
Group 2 N=43	34.524	16.286	33.405	84.143
t-value	.799	.049	1.068	.848
d.f.	62	62	62	62

Table 3 indicates no significant differences in delayed post-instructional observations for between Group 1 and Group 2.

TABLE 4. t-TEST COMPARISON OF POST-INSTRUCTIONAL AND DELAYED POST INSTRUCTIONAL ATTITUDE SCORE MEANS FOR GROUP 2

	Subscore variables			
	Evaluative	Potency	Oriented activity	Total
Post	34.150	16.550	32.925	83.575
Delayed Post	34.250	16.200	33.125	83.550
t-value N=40	.174	1.125	.269	.053

Table 4 indicates that there was no significant difference between post and delayed post-instructional observations for Group 2.

TABLE 5. t-TEST COMPARISON OF PRE-INSTRUCTIONAL AND POST-INSTRUCTIONAL MEAN ATTITUDE SCORES

	Subscore variables			
	Evaluative	Potency	Oriented Activity	Total
Pre-test	19.414	9.966	19.897	49.207
Post-test	24.966	11.931	24.793	61.655
N	29	29	29	29
d.f.	28	28	28	28
t-value	6.009***	4.309***	4.420***	5.343***

***$p < 0.001$

Table 5 indicates that there were significant differences between pre-instructional and post-instructional observations on attitudes. An examination of the means revealed that post-instructional scores were higher.

Question 3. Do the subjects' attitudes toward the Taxonomy affect their test construction behavior?

In answering this question, the relationship between the delayed post-instructional attitude scores and proportions of knowledge level items on sample test 2 were examined. The

correlation coefficient was calculated using the cosine-pi approximation for dichotomous data.

TABLE 6. CALCULATION OF TETRACHORIC CORRELATION COEFFICIENT BETWEEN DELAYED POST-INSTRUCTIONAL ATTITUDE SCORES AND PROPORTION OF KNOWLEDGE LEVEL ITEMS ON SAMPLE TEST 2

		Proportion of Knowledge Level Items	
		Number of subjects above median	Number of subjects below median
Attitude scores	above median	14	18
	below median	18	13

r_t = -.210 d.f. = 61

Table 6 indicates a negative relationship between proportion of knowledge level items and attitude scores, however, the relationship was not statistically significant..

Question 4. Does the subjects' test-construction behavior differ from the test-construction behavior of a population of science teachers?

To answer this question, four sets of test construction data were compiled into a summary table. Set A consists of sample test 1 for all science methods students, set B consists of sample test 1 for the study population, set C consists of sample test 2 for the study population, and set D consists of tests prepared by a population of science teachers reported in a NARST paper presented by Jacobson (1968).

TABLE 7. PROPORTIONS OF ITEMS OCCURRING AT THE SIX LEVELS OF THE TAXONOMY FOR SAMPLE TEST 1 FOR ALL METHODS STUDEN (A): SAMPLE TEST 1 FOR THE STUDY POPULATION (B): SAMPLE TEST 2 FOR THE STUDY POPULATION (C): AND TESTS PREPARED BY VIRGINIA SCIENCE TEACHERS (D)

	Major	Cognitive levels					
		1	2	3	4	5	6
	General science	54.13	28.15	12.83	1.53	1.05	2.33
A	Biology	54.25	26.47	11.16	3.57	2.40	2.16
	Physical science	25.29	28.12	40.06	2.94	1.59	2.00
	General science	52.90	30.05	12.62	1.19	1.14	2.10
B	Biology	52.07	26.85	11.15	3.23	3.45	3.25
	Physical science	34.33	24.50	37.00	1.50	1.00	1.67
	General science	68.52	20.10	8.14	.71	1.05	1.48
C	Biology	74.05	19.20	4.93	.75	.18	.90
	Physical science	34.83	18.50	43.67	.83	1.67	.50
	General science	85.55	8.79	5.65	0	0	0
D	Biology	84.71	13.96	.95	.30	.05	0
	Physical science	64.71	16.13	19.04	.02	.10	0

Cognitive levels:

1. knowledge
2. comprehension
3. application
4. analysis
5. synthesis
6. evaluation

Insufficient data were available to answer Question 4 statistically. An examination of Table 7 shows that the tests prepared by the study population both in the methods course and during the student teaching experience contained a lower proportion of knowledge level items than the tests prepared by a population of science teachers reported in an earlier study.

Summary and Conclusions

The data obtained from the taxonomic analysis of the tests indicates that the Taxonomy is a tool that can be used with reasonable precision to classify test questions. An inter-scorer reliability coefficient of .89 and in intra-scorer reliability coefficient of .94 were obtained.

The factorial analysis of the test data revealed:

1. no significant differences between the levels of grade point average.

2. that tests prepared by physical science majors contained significantly lower proportions of knowledge level items than tests prepared by biology and general science majors. There were no significant differences in proportions of knowledge level items on tests prepared by biology and general science majors.

3. that tests prepared by the study population during the student teaching assignment contained significantly higher proportions of knowledge level items than the tests they prepared during the methods course.

When the test data from the study population were compared with test data from a population of science teachers, it was found that both sets of tests prepared by the study population contained lower proportions of knowledge level items than tests prepared by a population of science teachers.

The attitude data indicated that instruction caused a significant positive change in attitude that persisted through the student teaching assignmetn.

When the relationship between proportions of knowledge level items and attitude data were examined, a negative correlation was observed; however, the coefficient was not statistically significant.

In view of the findings of the study, it seems appropriate to conclude that cognitive instruction with pre-service secondary science teachers can increase the cognitive levels of items on the tests they prepare. The findings further indicate that cognitive instruction in the use of the Taxonomy causes a positive shift in attitudes that appears to persist.

It is recommended that further research be conducted to determine the variables affecting the test-construction behavior of teachers. Studies should be conducted to determine the efficacy of different levels of questions in producing desired cognitive and affective changes in learner behavior.

BIBLIOGRAPHY

Andersen, Hans O. Toward More Effective Science Instruction in the Secondary School. New York: Macmillan, 1972.

Baker, Eva L. Relationship between learner achievement and instructional principles stressed during teacher preparation. Journal of Educational Research, 1969, 63(3), 99-102.

Bloom, Benjamin S. (Ed.). Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: David McKay, 1956.

Bloom, Benjamin S., Hastings, J. Thomas, and Madaus, George F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

Cox, Richard C., and Wildemann, Carol E. Taxonomy of Educational Objectives: Cognitive Domain, An Annotated Bibliography. Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pa., 1970.

Dahlberg, John Eric Jr. An Analysis of the Relationship Between the Cognitive Level of Teacher Questions and Selected Variables. (Doctoral Dissertation, University of Oregon). Dissertation Abstracts, 1970, 30 (7-8) 3344-3345A.

Davis, O.L. Jr., & Hunkins, Francis P. Analysis of textbook questions. Peabody Journal of Education, 1966, 43, 285-292.

Gagne, Robert M. The Conditions of Learning. (2nd ed.) New York: Holt, Rinehart and Winston, 1970.

Hunkins, Francis P. Bloom's taxonomy as a test construction guide. Ideas Educational, The Kent State University School, 1966, 4(2), 13-16.

Hunkins, Francis P. The influence of analysis and evaluation questions on critical thinking and achievement in sixth grade social studies. Educational Leadership, 1968, 25, 326-332.

Jacobson, Milton. The relationship of attitudes, self-concept, intellectual factors and teaching experience. Paper presented at the meeting of the Forty-first National Association for Research in Science Teaching, Chicago 1968.

Krathwohl, David R. et al. Taxonomy of Educational Objectives Handbook II: Affective Domain. New York: David McKay, 1964.

Kropp, R. P., Stoker, H. W., & Bashaw, W. L. The Construction and Validation of Tests of the Cognitive Process as Described in the "Taxonomy of Educational Objectives." Cooperative Research Project No. 2117, U.S. Office of Education, Institute of Human Learning, and Department of Educational Research and Testing, Florida State University, Tallahassee, 1966.

Ladd, George T., and Andersen, Hans O. Determining the level of inquiry in teachers' questions. Journal of Research in Science Teaching, 1970, 7(4), 395-400.

Osgood, C.E., Suci, G. J., & Tannenbaum, P.H. The Measurement of Meaning. Urbana: University of Illinois Press, 1957.

Pfeiffer, Isabel & Davis, O.L. Jr. Teacher-made examinations--what kind of thinking do they demand? Bulletin of the National Association of Secondary School Principals. 1965, 49(302), 1-10.

Shulman, Lee S. "Psychology" in Mathematics Education, Sixty-ninth Yearbook of the National Society for the Study of Education, 1970, 64.

Stevens, Romiett. The question as a measure of efficiency in instruction. Unpublished Doctoral Dissertation, New York: Teachers College, Columbia University, Contribution to Education, No. 48, 1-95, 1912.

Talley, Lawrence H., and Solomon, Gerard O. A study of cognitive behaviors of a college level faculty. Paper presented at the meeting of the Forty-fourth National Association for Research in Science Teaching, Silver Spring, March, 1971.

Tyler, Ralph W. What high-school pupils forget. Ohio State University Educational Research Bulletin, 1930, 9, 490-492.

Williams, Tony L. The effect of cognitive instructions on secondary student teachers and their pupils. Journal of Research and Development in Education, 1970, 4, 73-83.

WORD ASSOCIATION SCALE

Directions: The purpose of this form is to measure the meaning of certain words to persons acquainted with Bloom's Taxonomy. The form contains a variety of words to be rated. On the following page beneath the concept "Bloom's Taxonomy" is a series of rating scales. You are to rate that concept on each of the scales in order. Be sure to make your judgments on the basis of what the concept means to you, and mark the scale honestly.

Here is how the scale is to be used:

If you think the concept is very closely related to one end of the scale, you should mark the scale as follows:

sweet__X__:_____:_____:_____:_____:_____:_____sour

or

sweet_____:_____:_____:_____:_____:_____:__XX__sour

If you think the concept is quite closely related to either end of the scale, you should mark the scale as follows:

smooth_____:__X__:_____:_____:_____:_____:_____rough

or

smooth_____:_____:_____:_____:_____:__X__:_____rough

If the concept seems only slightly related to one end of the scale, mark it as follows:

long_____:_____:__X__:_____:_____:_____:_____short

or

long_____:_____:_____:_____:__X__:_____:_____short

The direction toward which you mark depends upon which end of the scale seems most characteristic of the concept you are judging.

If you consider the word to be neutral on the scale, or if the word is not related to either end of the scale, or if the concept is equally related to both words on the scale, then mark as follows:

wet_____:_____:_____:__X__:_____:_____:_____dry

Be sure to place your marks in the blanks, not on the (:).

Be sure to check every scale for it is important that you omit none; but never put more than one mark on each scale. Make each an independent judgment.

Name_____

Date_____

Rank the concept "Bloom's Taxonomy" on each of the following scales.

bad _____:_____:_____:_____:_____:_____:_____good

strong _____:_____:_____:_____:_____:_____:_____weak

relevant _____:_____:_____:_____:_____:_____:_____irrelevant

impotent _____:_____:_____:_____:_____:_____:_____potent

shallow _____:_____:_____:_____:_____:_____:_____deep

successful _____:_____:_____:_____:_____:_____:_____unsuccessful

passive _____:_____:_____:_____:_____:_____:_____active

regressive _____:_____:_____:_____:_____:_____:_____progressive

reputable _____:_____:_____:_____:_____:_____:_____disreputable

interesting_____:_____:_____:_____:_____:_____:_____uninteresting

obstructive_____:_____:_____:_____:_____:_____:_____helpful

valuable _____:_____:_____:_____:_____:_____:_____worthless

applicable _____:_____:_____:_____:_____:_____:_____unapplicable

positive _____:_____:_____:_____:_____:_____:_____negative

important _____:_____:_____:_____:_____:_____:_____unimportant

Additional comments: